UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Technology

Michael Joseph

The effect of binding sites for POU-HD proteins (Oct-1) on the replication of HPV18

Master's Thesis (30 ECTS)

Curriculum Bioengineering

Supervisor(s):
Associate professor, PhD Marko Piirsoo

PhD student, MSc Martin Kala

Tartu 2021

The effect of binding sites for POU-HD proteins (Oct-1) on the replication of HPV18

Abstract: Human papillomaviruses (HPV) are known to infect the basal cells of the epithelia. Infection caused by certain type of HPVs are known to lead to cancer. Therefore, understanding the viral life cycle could be a key to develop novel therapeutic strategies to inhibit viral infection. The regulation of HPV genome replication depends upon the certain viral factors and various cellular factors which could potentially have a positive or a negative effect on the viral replication. The upstream regulatory region (URR) of the HPV genome has motifs for these viral and cellular factors. One such cellular factor, Oct-1 has multiple binding sites located in this region. Based on a previous study it has been shown that Oct-1 binding does have a negative effect on the HPV18 minimal origin. Since, the URR could be filled with a lot of potential motifs for Oct-1 factors, it is important to understand the effect of these specific motifs on the HPV genome replication. This thesis aims to observe the impact of such binding sites on HPV18 replication by performing mutational analysis.

Keywords: Human papillomavirus (HPV), viral genome replication, cellular factors, Oct-1 transcriptional factor.

CERCS: B230 Microbiology, bacteriology, virology, mycology.

POU-HD transkriptsioonifaktorite sidumissaitide mõju HPV18 replikatsioonile

Lühikokkuvõte: Inimese papilloomiviirused (HPV) nakatavad mitmekihilise epiteeli basaalseid rakke. Teatud HPV tüüpide nakkus võib viia vähitekkele. Seega on HPV elutsükli mõistmine oluliseks teguriks uute viirusvastaste strateegiate väljatöötamisel. HPV genoomi replikatsioon nakatunud rakus sõltub nii viiruslikest kui ka rakulistest faktoritest, mis võivad omada nii positiivset kui ka negatiivset mõju. HPV genoomi mittekodeeriv osa (URR) sisaldab sidumissaite paljudele rakulistele valkudele sealhulgas mitmeid saite transkriptsioonifaktorile Oct-1. Eelnevad uuringud on näidanud, et Oct-1 omab negatiivset mõju onkogeense HPV18 URR-i replikatsioonile. Kuna Oct-1 seondumiskohti on URR-is mitmeid, on oluline aru saada nende individuaalsest mõjust viirusgenoomi replikatsioonile. Käesoleva töö eesmärgiks on uurida Oct-1 seondumissaitide mõju HPV18 genoomi replikatsioonile

Võtmesõnad: Inimese papilloomiviirus (HPV), viirusgenoomi replikatsioon, rakulised faktorid, Oct-1 transkriptsioonifaktor.

CERCS: B230 mikrobioloogia, bakterioloogia, viroloogia, mükoloogia.

TABLE OF CONTENTS

TERMS, ABBREVIATIONS AND NOTATIONS

AMF-1 – Autocrine mobility factor 1

AP-1 – Activator protein

Bp – Base pairs

Brd4 - Bromodomain-containing protein

CBP – CREB binding protein

cEBP - CCAAT-enhancer-binding proteins

Ct – Cycle threshold

FACS - Fluorescence-Activated Cell Sorting

GSP2 – G protein pathway suppressor 2

HPV - Human papillomavirus

HSPG - Heparin sulphate proteoglycans

LB – Lysogeny broth

MC - Minicircle

NAP-1 – Nucleosome assembly protein

Nluc – Nano luciferase

ORF – Open reading frame

P/CAF - CBP/p300-binding protein

pRb – Retinoblastoma protein

PV – Papillomaviruses

RLU – Relative Luciferase Units

RPM – Rotation per minute

SB – Southern Blotting

SP1 - Specificity protein 1

TB – Terrific broth

TBP- TATA binding protein

TFIID – Transcription factor II D

TFIIB - Transcription factor II B

TopB1 - DNA Topoisomerase II Binding Protein 1

URR – Upstream regulatory region

WT- Wild type

INTRODUCTION

Papillomaviruses that infect humans and animals make up a broad family of viruses consisting of different subtypes. Human papillomavirus (HPV) infections cause warts and lesions and depending on the site of infection these lesions have an increased risk of developing into tumours. Such cancer inducing HPVs are categorized as high-risk types. HPVs are the primary cause for all cases of cervical cancers and are responsible for 5.2% of all malignancies worldwide. Vaccines have been developed to prevent the infections, but no drugs or therapeutics exist to cure ongoing infections.

Studies have been done and are currently ongoing to understand the HPVs viral life cycle and its interactions with the infected host cells. When the virus infects the epithelium, specifically the basal layer of the epithelium, the replication of viral genome will be initiated, and viral proteins will be expressed. Various viral proteins and host cellular factors come into play throughout the infection cycle, which will end in the assembly of virions. It is important to understand the molecular mechanisms involved during HPV infections, to develop therapeutic strategies which could inhibit the viral replication. The host cell has different cellular factors which could recognize certain motifs along the upstream regulatory region (URR) of the viral genome. Although, various papers have been published in order to help us understand which sequence specific cellular factors are involved in HPV infection, yet a lot is unknown of their potential effect on HPV genome replication.

This thesis involves the understanding of one such sequence specific (octamer sequence) cellular factor and its potential effect on HPV replication. Transcription factor Oct-1, a member of the POU-homeodomain family of transcriptional factors, could bind to the octamer sequences and related sequences (binding sites) on the URR. The aim of this study is to investigate the effect of such binding sites on HPV genome replication, by performing mutational analysis.

1 LITERATURE REVIEW

1.1 Papillomaviruses

Papillomaviruses (PV) come under the *Papillomaviridae* family. These viruses have dsDNA genome around the size of 8 kilobases, which has 8 open reading frames (ORF). The DNA is contained inside an icosahedral capsid. Papillomaviruses are epitheliotrophic, they infect keratinocytes of the cutaneous and mucosal epithelium. Papillomaviruses are known to infect a variety of hosts such as reptiles, mammals, and birds (de Villiers et al., 2004). Categorization of different types of papillomaviruses is based on the viral L1 gene sequence, which is the most conserved region in different types of PV genomes. The Papillomaviruses are divided into two subfamilies, mucosal and cutaneous, both including over 50 genera and more than 300 different species (van Doorslaer et al., 2018). Based on the type of PV and the site of infection, they cause warts or known as papillomas, benign tumours, and may even lead to the development of cancer.

1.2 Human Papillomaviruses

Human papillomaviruses (HPV) are a diverse group of papillomaviruses which infect the human epithelial tissues. They are divided into 5 groups namely alpha - (α), beta- (β), gamma- (γ), mu- (μ), and nupapillomaviruses (ν); more than 200 distinct types of HPV genomes have been sequenced (Doorbar et al., 2012) (Figure 1). Each group is distinct in terms of infection cycle and disease associations. Furthermore, they are also categorized as high-risk and low-risk types. The term high-risk is based on the ability of the virus to expose the infected person to the risk of developing various malignancies.

The World Health Organization (WHO) has labelled a dozen HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) as high-risk cancer-causing types and among these HPV 16 and 18 are found to be most prevalent in different cancer specimens (Doorbar et al., 2012).

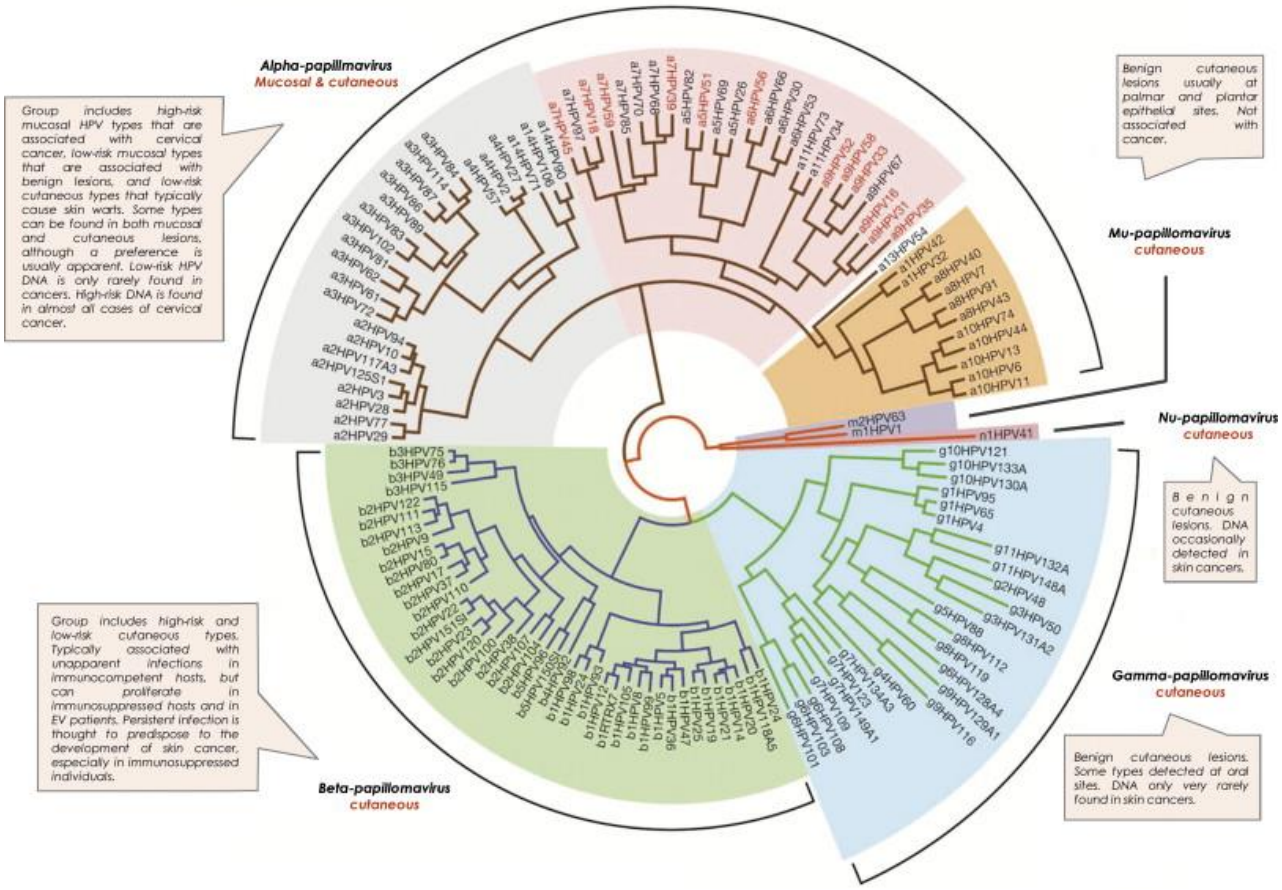

Figure 1: Chart depicting the different groups of human papillomaviruses (HPV) (Doorbar et al., 2012).

1.3 HPV Infection cycle

HPV has a specificity for the basal layer of the stratified squamous epithelium, mainly because the basal cells are retained in the cell cycle and dividing into daughter cells, hence HPV infection cycle is connected to the differentiation of the basal layer of the stratified squamous epithelium, present as the epidermal layer of the skin and also as the mucosal layer of vagina and anal canal.

Through micro-abrasions or injuries, the basal layer of the epithelium could be exposed. HPV infects the cells of the basal layer of epithelium through cellular receptors (Malik et al., 2014). Depending on the type of HPV and cell type it infects, different receptor strategies or combination of strategies could come into play during the infection (Graham, 2017b). The viral capsid protein L1 binds to the cellular receptors present on the membrane of the basal layers. Heparin sulphate proteoglycans (HSPGs) are thought to be the primary receptors

HPVs bind to (Graham, 2017b), (Malik et al., 2014). Most HPVs enter the cell through clathrin-dependent endocytosis and while some HPV types enter the cell through caveolae-dependent endocytosis (Malik et al., 2014).

While internalized, the virions are unpacked; L1 protein remains in the endosome while the L2 protein-viral genome complex facilitates the nuclear transport of the genome and establishes itself as episome inside the nucleus (Doorbar et al., 2012). This brings about the initial stage of viral infection, where viral gene expression is activated, early proteins are expressed, and viral genome replication starts. The viral genome copy number reaches about 50-500 per cell during the initial infection. In the next stage, as the basal cells divide, the viral genome number is doubled with the rate of host cell DNA synthesis and equally divided during the mitosis among the daughter cells, hence maintaining the viral genome copy number. The last stage involves the vegetative replication of the viral genome, where genome copy number rises again, the differentiated keratinocytes express late proteins for the assembly of virions and new HPV viral particles are released as the cells reach the epithelial surface (Kadaja et al., 2009).

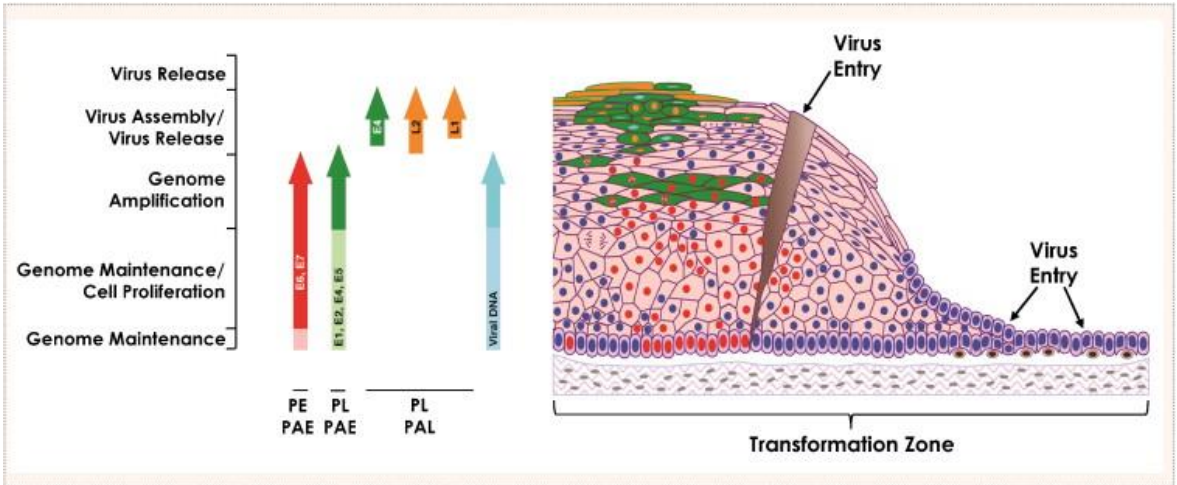

Figure 2: General overview of the infection cycle of HPVs in the epithelial cells (Doorbar et al., 2012)

1.4 HPV Genome

All HPV genomes have a similar structure, with a circular double-stranded DNA genome of around 8000 base pairs and have 8 open reading frames (ORFs) (Burley et al., 2019) (van Doorslaer et al., 2017) (Figure 3). 6 are situated in the "early" region and 2 in the "late" region. Early proteins are involved in the HPV genome replication and transcription, modulation of the host cell cycle, cell signalling, apoptosis, modulation of the hosts immune response, and structural modification of infected cells, among other things (Graham, 2017a). These early proteins are expressed throughout the infection cycle. The late proteins are expressed in the later stage of the infection cycle, these proteins are required for the viral capsid formation and virion transmission (Graham, 2017a). Upstream to these open reading frames is a non-coding region known as the upstream regulatory region (also as the long control region). This regulatory region contains various cis-enhancer elements and binding sites to a lot of cellular and viral factors which influence the viral replication and transcription (Bromberg-White & Meyers, 2002).

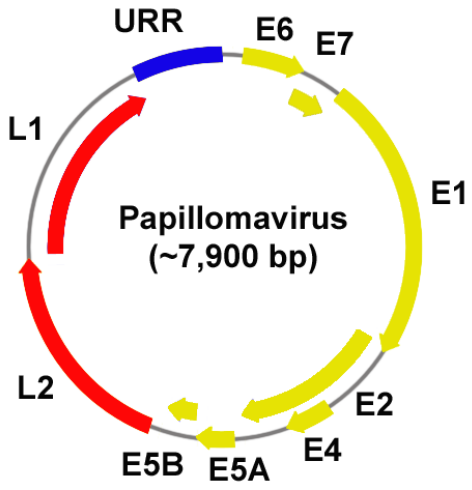

Figure 3: Schematic representation of the HPV genome which is divided into the upper regulatory region (URR), the early region and the late region.

1.5 URR

The “early” and the “late” regions are connected by a ~850 base pair long non-coding region which contains the origin of replication (Ori) and binding sites (cis-enhancer elements) for different transcriptional factors (Wang et al., 2011). These factors regulate the replication and gene expression in HPV.

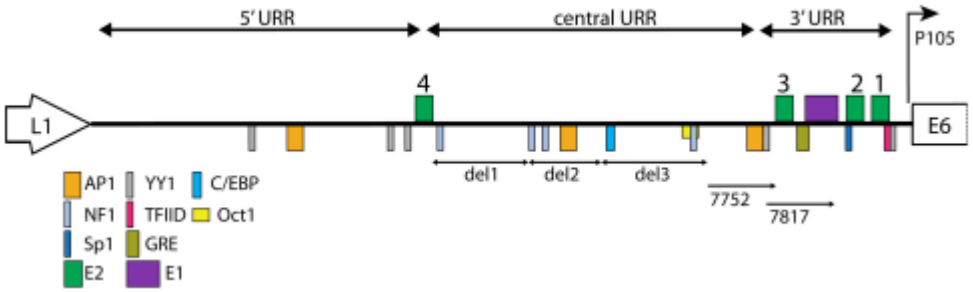

Figure 4: The HPV 18 URR with an arrangement of binding sites for viral and cellular factors (van Doorslaer et al., 2017). The URR has a lot of known and putative binding sites for different cellular factors.

Most HPVs have a similar arrangement of cis-enhancer elements, especially the high-risk types 16 and 18 (Stu¨nkel & Bernard, 1999). All HPVs have one origin of replication (Ori), one E1 binding site and more than one E2 binding sites (Sverdrup & Khan, 1995). HPV 18 has four E2 binding sites and the URR is generally divided into 3 segments namely the 5’ end, the central region and the 3’ end (Figure 4) (van Doorslaer et al., 2017). E2, along with E1, plays a role in viral replication where they have binding sites on the origin of replication (Sverdrup & Khan, 1994). E2 also acts as a viral transcriptional regulator by binding to other E2 specific sites on the URR (Demeret et al., 1995).

1.5.1 Early Region

The early region of all different types of HPVs is located at 5’ end of the genome and has a few common open reading frames for early proteins such as E1, E2, E4, E5, E6, E7 and E8 (Wang et al., 2011). The core proteins E1 and E2 are necessary for viral DNA replication and amplification, as well as viral transcription control (Burley et al., 2019). Proteins from the other ORFs such as E4, E5, E6, and E7 are involved in interactions with the host cell,

which help to create an environment suitable for virus to replicate in the keratinocytes (Burley et al., 2019). Most early proteins are expressed throughout the infection cycle at different expression levels, but E4 is expressed towards the later stage, and is involved in cytokeratin network remodelling, cell cycle arrest and virion assembly. E5 is associated with control of cell growth and differentiation, E6 is known to be the oncoprotein as it inhibits apoptosis and cellular differentiation. E7 controls the cell cycle and centrosome duplication. (Graham, 2010).

1.5.2 Late Region

Downstream to the early region lies the late region that encodes for viral capsid proteins L1 and L2 (Wang et al., 2011). As the name suggests this region is expressed during the later phase of the infection in terminally differentiated keratinocytes, during which assembly of virions takes place (Wang et al., 2011). L1 is the major capsid protein which forms the capsid enveloping the L2-genome complex and is recruited by minor capsid protein L2, which plays a role in delivery of the viral genome into the infected cells and viral genome encapsidation (Burley et al., 2019), (Graham, 2010).

1.6 HPV and cancer

Most commonly HPV infection causes lesions and warts, and usually the infection goes away by itself. But continuous persistent infections have a high chance of developing into cancer (Doorbar et al., 2012). High-risk HPV type infections could last for a long time, sometimes for several years, and drive cell proliferation in the basal and suprabasal cell layers at specific sites of infection, the exact reason for this is yet to be understood completely. It has been shown that increased expression of the viral proteins E6 and E7 are associated with the risk of developing tumors (Doorbar et al., 2012), (Burley et al., 2019), (Kitagawa et al., 1996). In immature epithelial cells, the viral oncoproteins E6 and E7 disrupt the activity of p53 and pRb oncosupressors, which leads the epithelial cells of the cervix to immortalise, and this immortalisation leads to the development of cervical cancer (Feng et al., 2020).

Cervical cancer is a common cancer among women and more than 99% of all cases of cervical cancer are caused by HPV infection (Liu et al., 2015). High-risk HPVs 16, 18, 31, and

33 are the main types associated with cervical cancer (Williams et al., 2011). HPV is also known to cause cancer to other regions of the body such as vagina, penis, anus, and in some cases to the oropharynx (Wang et al., 2011), (Gillison et al., 2008), (Saraiya et al., 2015). Out of all the cancer diagnosed world-wide, HPV related cancers make up to 5% of the total cancer cases; statistically estimated to be causing half a million cases per year (Parkin, 2006).

1.7 HPV replication mechanisms

After the virus has gained access to the basal layer of the epithelium through micro-abrasions and the viral genome has reached the nucleus of the basal cell, the initial stage of infection starts with the expression of early proteins E1 and E2, guiding the host cell replication machinery and triggering the initial amplification of the viral genome (Burley et al., 2019), (Orav et al., 2015). The copy number of the viral genome reaches about 50-500 per cell (Figure 5) (Fisher, 2015).

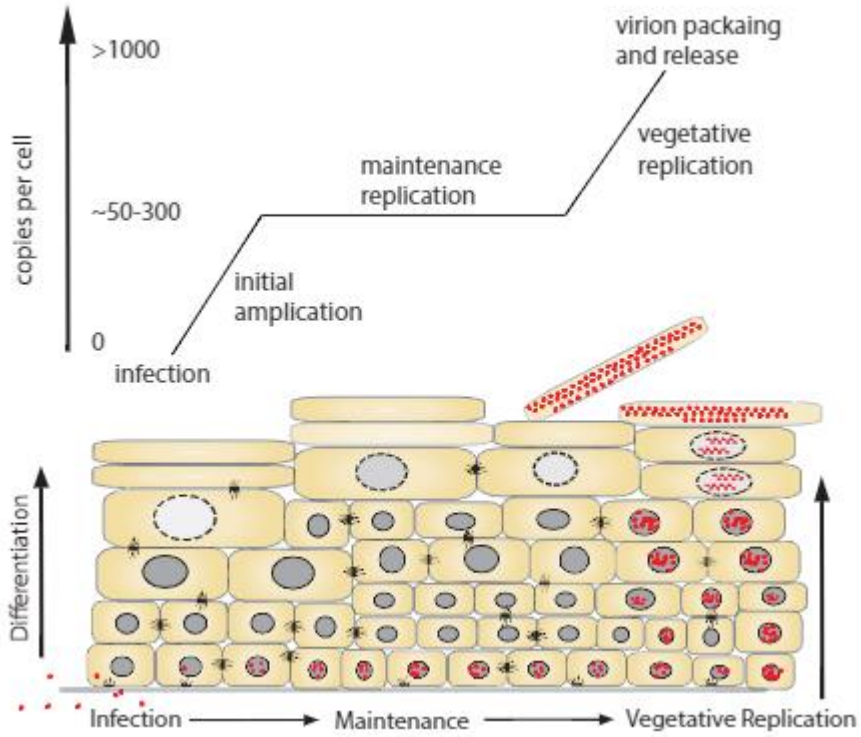

Figure 5: Different stages of HPV replication (Fisher, 2015). The viral genome starts replicating as soon it reaches the nucleus of the cell and goes through different phases of replication during the infection cycle.

To maintain the copy number in dividing daughter cells, the viral genome is doubled during the S phase of the cell cycle and is equally divided into the two daughter cells, this is known as the maintenance replication (Hoffmann et al., 2006), (Kadaja et al., 2009). It is suggested that there are two alternative modes by which the viral genome is replicated and maintained at a stable copy number. First mode is when the replication of the viral genome happens once per S phase, along with the cellular DNA. Second is by random replication where the genome may replicate a few times once per cycle or one time per cycle or may not replicate at all. In theory both these modes maintain the copy number per cell at a constant range in differentiating keratinocytes (Kadaja et al., 2009) (Hoffmann et al., 2006). E2 plays a role in equal segregation of viral genome among daughter cells by anchoring its motifs on the URR to the host cell chromosome (Kadaja et al., 2009).

As the daughter cells move upwards and are differentiated, the vegetative phase replication begins, which involves the continuous amplification of the viral genome resulting in an increase in copy number to around 1000 per cell (Fisher, 2015). The expression of late proteins L1 and L2 facilitates the encapsidation of the viral genome and hence resulting in the assembly of virions. Once the differentiated cells reach the epithelial surface the virions are released (Moody, 2017).

1.8 Cellular factors in HPV replication

Different stages of HPV infection cycle are dependent on certain cellular factors and their interaction with the viral genome and proteins. The host cellular replication machinery such as DNA polymerase α/primase, replication protein A, single stranded DNA binding protein RPA and topoisomerase I along with E1 and E2 proteins assemble at the origin of replication during the replication of the HPV genome (Burley et al., 2019), (Archambault, 2011). The vast number of cis-responsive elements or motifs for various cellular factors on the URR plays a role in HPV genome replication and transcription regulation. These specific sequences often differ between different types of HPV. The spacing between SP1, TATA and E2 binding sites is mostly conserved in all mucosal HPVs such as HPV 16 and 18, the overlapping of E2 binding sites with these cis-elements and the interplay between SP1 and TFIID and E2 viral protein brings about the regulation of HPV transcription by activating and inactivating the early promoters (Ribeiro et al., 2018), (Archambault, 2011). The chromatin

associated factor, bromodomain-containing protein (Brd4) acts as the interacting protein for E2 viral protein for chromosome tethering (Archambault, 2011). E2 is known to associate with a bunch of different cellular and transcriptional factors (Archambault, 2011) such as TFIID, TFIIB, TBP (Miller Rank & Lambert, 1995), AMF-1/GPS2 (Breiding et al., 1997), p300/CBP (Lee et al., 2000), NAP-1 (Rehtanz et al., 2004), P/CAF (Lee et al., 2002), TopB1 (Boner et al., 2002).

Typical HPV URR contains a cluster of E2 binding sites and certain host cellular factors may have one or multiple binding for example AP-1 (Chan et al., 1990), cEBP (Bauknecht et al., 1996), NF1 (Chong et al., 1990), NF-IL9 (Kyo et al., 1993), Oct-1 (O'Connor & Bernard, 1995), SP1 (Gloss & Bernard, 1990), YY1 (Dong et al., 1994). Transcription factors such as AP-1, NF1, Oct-1 and SP1 have exhibited the ability to activate transcription from HPV 16 early promoter p97, while YY1 has shown both activation and repression of transcription from the early promoter (Carson & Khan, 2006).

Certain transcriptional factors such as c-Myb, C/EBPα, C/EBPβ, NFAT, YY1, NF1, Oct-1, c-Jun and SP1 have been shown to bind to late promoter region (Moody, 2017). HPV transcript splicing is positively regulated by Serine Arginine splicing factors (SRSF) and by CTCF insulator proteins which has a binding site on E2 ORF of the viral genome. This site is also known to be well conserved among high-risk HPV types (Moody, 2017).

Quite a few host cellular factors such as CyP and PyK2 have been shown to be involved in viral capsid disassembly and separation of L1 protein L2 protein genome-complex, during the viral infection (Aksoy et al., 2017). Besides regular cellular factors, certain glucocorticoid response elements (GREs) also have been found to have binding sites on the URR of quite a few HPV types such as HPV 11, HPV 16, and HPV 18 (Bromberg-White & Meyers, 2002).

Protein kinases such as casein kinases 1 and 2, cyclin dependant kinase, FGF receptor 3 protein kinase A and C and MAP kinase, have been shown to regulate E1 and/or E2 activities and their cellular localizations (Piirsoo et al., 2019). E1 phosphorylation by cyclin/Cdk complexes is suggested to play a role in the initial stages of replication of HPV (Ma et al., 1999). Activation of E1 nuclear localization sites by Mitogen-activated protein kinases is necessary for E1 to transfer into the nucleus for DNA replication, and to establish a persistent HPV infection (Yu et al., 2007). CK2 is shown to be required for transient replication of different

HPV types, certain inhibitor molecules have shown to target CK2 activity which leads to degradation of E1 and hence hindering HPV DNA replication (Piirsoo et al., 2019).

1.9 POU factors

The name POU is derived from the primary members of the POU family, Pit-1, Oct-1, Oct-2, and Unc-86. The POU factors belong to a family of DNA-binding transcription factors that play a role in cell type specification and development (Gold et al., 2014). POU family of transcription factors are defined by the common 150-160 amino acid domain (Latchman, 1999). POU factors have been identified to have a N-terminal POU-specific domain and a C-terminal POU-homeodomain which is connected by a linker region. (Gold et al., 2014), (Phillips & Luisi, 2000).

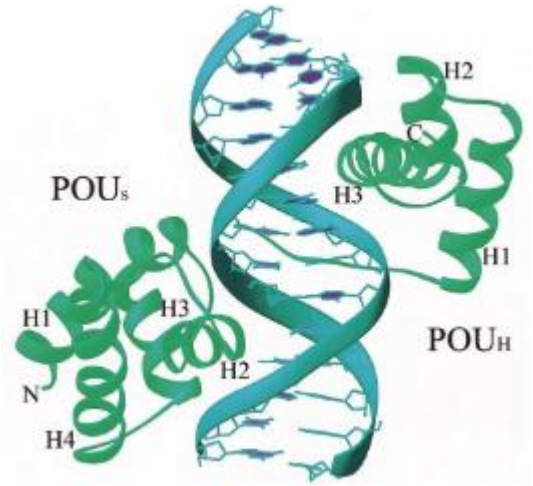

Figure 6: Crystal Structure of Oct-1 POU-HD complexed with DNA (Phillips & Luisi, 2000).

Members of the POU family recognize specific DNA sequences (cis-responsive elements) (Phillips & Luisi, 2000). For example, Oct-1 binds to a specific eight-base pair DNA sequence and similar variants of that sequence (Phillips & Luisi, 2000).

These sub-domains bring about the interaction of these cellular factors to specific DNA sequences in the regulatory region of many genes (Latchman, 1999). The α-helices of subdomains, POU-specific and POU-homeodomain attach to the DNA through hydrogen bonds (Gold et al., 2014), (Reményi et al., 2001). POU factors can undergo conformational changes when they interact with DNA sequences and other transcriptional factors, they form heterodimers and homodimers that bind to DNA (Gold et al., 2014). Some POU proteins can have

different isoforms which could either compete with other transcriptional factors or work together by binding to multiple transcriptional factors (Konzak & Moore, 1992). Minor differences in spacing between the two DNA binding domains of the protein can make it act as a repressor in one situation and an activator in another (Gold et al., 2014), (Scully et al., 2000).

1.10 Oct-1 and its effect on HPV replication

Oct transcription factors of the POU family proteins bind to an 8 bp DNA sequence (ATGCAAAT) known as the octamer consensus motif (Sturm et al., 1988). There are 13 POU transcription factors in humans and mice, with eight of them displaying specificity for the octamer consensus motif and hence being called Oct proteins. Oct-1 is the only octamer binding (Oct) protein that is widely expressed among different organs. Compared to other Oct proteins, Oct-1 is the most prevalent in various epithelial tissues. (Vázquez-Arreguín & Tantin, 2016). Oct-1 is also known to interact with other transcriptional factors and partner with itself to recognize binding sites and bring about regulation of transcription (Vázquez-Arreguín & Tantin, 2016).

Oct-1 has been found to be involved in regulation of the activity of the URR (Morris et al., 1993), (Latchman, 1999). HPV URR have a few binding sites (octamer motifs) for Oct-1 to recognize and bind to. Even a few degenerate versions of the octamer motif can be recognized by Oct-1 (O'connor et al., 1995). The presence of such motifs along the URR is yet to be fully scoped and the potential effect of such sites on the regulation of HPV replication is still unknown.

2 THE AIMS OF THE THESIS

The aim of this thesis is to perform mutational analysis of the identified and putative Oct-1 binding sites on the HPV18 URR to understand their role in the replication and transcription of HPV18 genomes.

2.2 Objectives

- Generating individual mutant genomes by replacing predicted Oct-1 binding sites (octamer) to non-binding sequences.
- Transfect these mutants into U2OS cells and observe the effect of each mutation using luciferase and Southern blot assays.
- Investigate the impact of the given mutants have on the HPV18 transcription.
- Identify mutants which have the largest effect on HPV 18 replication and transcription in comparison to the wild type.

.

3. EXPERIMENTAL PART

3.1 Materials and methods

3.1.1 Designing insert sequences for cloning

Five potential Oct-1 binding sites in the URR of HPV18 and its vicinity were chosen based on previous publications and data gathered by our research group (unpublished) to be mutated. For that, six synthetic 1140 bp long sequences, overlapping the URR of HPV 18, which contained the mutated octamer binding sites, were ordered (Twist Bioscience) and cloned into pMC-HPV18 and pMC-HPV18Nluc vectors by replacing the native sequence. Five of the constructed mutants were designed to have only a single octamer site swapped with a nonbinding sequence, and the sixth mutant was designed to have all selected binding sites changed. The synthetic sequences were first cloned into the pJET 1.2/blunt vector using CloneJET PCR cloning kit (*Thermo Fisher Scientific*), according to the manufacturer's protocols to ensure unlimited amounts of the ordered sequences.

3.1.2 Cloning of the mutants.

For cloning of the synthetic insert sequences, pMC-HPV18 and pMC-HPV18Nluc (used for Nluc assays) plasmids were used as backbones. These plasmids contain full length HPV18 genomes. Nluc plasmids have additional nano-luciferase reporter gene, which could be detected to measure the replication efficiency of the viral genomes (Piirsoo et al., 2019). Both the pMC-HPV18, pMC-HPV18Nluc plasmids and the pJET 1.2 vector containing each of the 6 different insert sequences were digested using Bgl I and Bpu 1102 I enzymes and separated using 1.5% agarose-TAE gel. The inserts and the backbones were cut out from the gel and purified using Zymoclean Gel DNA Recovery Kit (*Zymo Research*) according to manufacturer's protocol. The digested inserts and backbones were also cleaned and concentrated using DNA Clean & Concentrator Kits (*Zymo Research*) according to the manufacturer's protocol.

After measuring the DNA concentration using Nano-drop spectrophotometer (*Thermo fisher*), the amount of insert and backbone required for ligation was calculated. For ligation, the insert to backbone ratio was kept 3:1 and reaction was done using 1 µl of T4 ligase incubated at 18°C overnight.

3.1.3 Transformation

Transformation was performed by heat-shock method. Competent *E. coli* DH5α cells were thawed on ice. 10 µl of ligation mix was added to 90 µl competent cells (1:10 ratio) and kept on ice for 30 minutes followed by heat shock at 42°C for 45 seconds and stored on ice for a minute. 800 µl of LB was added to pre-grow the culture at 37°C shaker for 1 hour. After pre-growth, LB plates containing Kanamycin antibiotic for selection were used to plate the samples, and then incubated overnight at 37°C.

After incubation, colonies were picked and grown in 3 ml LB containing Kanamycin (50 µg/ml) overnight at 37°C and 220 RPM shaker condition. The samples were then collected by centrifugation and plasmids were extracted using the FavorPrepTM Plasmid Extraction Kit (*Favorgen Biotech Corp*) according to the manufacturer's protocol.

3.1.4 Confirming the cloned mutants

The cloned plasmids were digested using the same enzymes that were used to clone the plasmids (Bgl I and Bpu 1102 I) and separated by gel electrophoresis to confirm successful ligation of the inserts. To further verify the generation of mutants, the plasmids were sent for sequencing (Sanger's sequencing). The sequences were aligned and compared with the original designed sequence as a reference to identify the correctly cloned plasmids.

3.1.5 WT HPV 18 and Nluc HPV 18 minicircle preparation

The mutant plasmids were converted to minicircles (MC), which has the functional HPV genome ready to be used for transfection. Cloned pMC-HPV18 and pMC-HPV18 Nluc plasmids were transformed into *E. coli* ZYCY10P3S2T strain (Toots et al., 2014). Transformation was done as mentioned above. The colonies from each plate were picked and grown overnight in 3ml LB with Kanamycin antibiotic. Next day, 200 µl of each overnight grown culture was added to individual flasks containing 100 ml of Terrific broth (TB) (*Difco*). TB contained 400 µl of glycerol and 100 µl of Kanamycin (50 µg/ ml) per 100 ml. The cultures were grown in 37°C shaker at 220RPM for 10-12 hours.

After 10-12 hours, 2 ml of culture was taken for plasmid extraction and restriction analysis. The induction mixture was added to the flasks to induce the removal of the pMC backbone by recombination and production of minicircles containing HPV18 genome. The induction

mixture was made from 100 ml LB, consisting of 0.04% of L-arabinose and 0.4M NaOH. The flasks were then incubated overnight at 32°C and 220 RPM in shaker.

Next day, another 2 ml of culture was taken from the flasks for plasmid extraction and restriction analysis to confirm the generation of MC's. The culture from the flask was transferred to centrifuge bottles and centrifuged at 10000 RPM and 4°C for 15 minutes. Supernatant was discarded and the cell pellets were used to extract minicircles by using Nucleobond® Xtra Maxi EF kit (*Macherey-Nagel*) according to the manufacturer's protocol. The extracted minicircles were re-precipitated using 1:10 volume of 5M NaCl and 2.5 volume of 96% ethanol to further purify extracted plasmids. Samples were vortexed and kept at -20°C for 30 mins and centrifuged at 4°C and 14000 RPM for 15 minutes, the samples were aspirated of the supernatant and the left-over cell pellets were washed with cold 70% ethanol followed by centrifugation again at 4°C and 14000 RPM for 5 minutes. The ethanol was aspirated, the cell pellets were left to dry for few minutes and then resuspended in nuclease free TE buffer.

The 2ml culture taken before and after induction were extracted for plasmids using Favor-Prep Plasmid Extraction Mini Kit (*Favorgen*), and then later restricted using Bgl I, BamH I and EcoR I and were run on gel electrophoresis. The restriction pattern helped to identify which cloned plasmids had formed minicircles.

3.1.6 Cell culture and Growth Media

U2OS cell lines have been shown to be a good system to study HPV infections (Geimanen et al., 2011). The cells were grown in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% fetal bovine serum (FBS), penicillin (100 IU/ml), and streptomycin (100 ng/ml) (*Sigma-Aldrich*). The cells were grown in 10 cm plates and incubated at 37°C and 5% CO_2 until they were confluent enough (about 90%) to be transfected.

3.1.7 Transfection of cells by electroporation

For all transfections U2OS cells had been split on the previous day. The media was aspirated, cells were washed with PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na2HPO4, and 1.47 mM KH2PO4). The PBS was aspirated, and cells were detached using 0.25% Trypsin-EDTA (*Sigma-Aldrich*). The detached cells were transferred to a 50ml falcon tube that had the same

volume of fresh media as the that of detached cells collected from every plate. The cells were centrifuged at 1000 RPM for 5 minutes. After centrifugation, the supernatant media was aspirated leaving the cell pellets behind. These cell pellets were resuspended with fresh media and along with Na-Bes (5 µM), the volume of fresh media added was calculated to hold ~10^6 cells per electroporation volume (250 µl).

The transfection mixtures were added to separate electroporation cuvettes (4 mm). The transfection mixture consisted of one of each individual HPV mutant genomes (1000 ng) plus the carrier 50 µg (Salmon sperm DNA). For some experiments 500 ng of pCDNA3 or 500 ng hOct1 plasmid was added for each electroporation depending on the sample. As a control, the wild type HPV18 genome was used along with the mutated ones. 250 µl of resuspended cells were added to this transfection mixture in the cuvettes. The mixture was electroporated using Gene Pulser XCell machine (*Bio-Rad instruments*) at 220V and 975 µF.

After electroporation 250 µl of fresh media was added to the cuvette and everything was transferred to tubes containing fresh media. The volume of media depended on the type of plates that were going to be used to seed the freshly electroporated cells. For southern blotting 6-well and for luciferase assay 96-well plates were used. The plates were also divided into different time points for which replication efficiencies were needed to be measured (48 hours and 96 hours). The plates were incubated at 37°C and 5% CO_2 and the media was replaced on the following day.

To check the transfection efficiency, plasmid expressing Green Fluorescence Protein (GFP) was transfected into the cells along with the HPV mutant genome samples. The transfection efficiency was measured by FACS analysis.

3.1.8 DNA extraction

To gather each time-point the cells were washed with PBS and 500 µl of Sol IV buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM EDTA, 0.2% SDS) was used to lyse the cells. Syringe and 23G needle were used to homogenize the lysate and collect them into tubes. 200µg/ml of Proteinase K was added to the tubes. The tubes were incubated at 56°C overnight. Next day, 500µl (1:1) of phenol-chloroform was added to the tubes, vortexed and centrifuged at 13000 RPM for 3 minutes. After centrifugation, the aqueous phase was carefully collected into new tubes. To this new tube 1 ml of 96% ethanol was added and vortexed. The tubes were kept at -20°C for 30 minutes, then spun down at 14000 RPM in 4°C for 15

minutes. The supernatant was aspirated carefully leaving the pellet behind. To this cold 70% ethanol was added to wash the pellet, the tubes were centrifuged again for 6 minutes. The supernatant was again aspirated carefully, and the tubes were left to dry for few minutes. TE buffer plus RNase A (10µg/ml) was added to the tubes and kept at 37°C for 1 hour. The DNA was reprecipitated using 5M NaCl and 2.5 volumes of 96% ethanol and again resuspended in TE buffer. The reprecipitated DNA from each sample was measured for its concentration using Nano-drop spectrophotometer.

3.1.9 Southern Blotting

For each sample, equal amount of DNA was digested. Genomic DNA samples were digested with 1 µl Dpn1 to cut the methylated DNA and 1 µl Bgl I to linearize the HPV 18 genome. Samples were loaded onto 0.8% agarose TAE gel for gel electrophoresis and separated overnight at 30mV

After the run, the gel was photographed on an UV transilluminator to confirm the DNA is loaded equally and treated with Solution A (0.5 M NaOH, 1.5 M NaCl) in a shaker for 30 minutes followed by treatment with Solution B (1 M Tris pH 7.4, 1.5 M NaCl) for 30 minutes. The DNA from the gel was transferred onto a nylon membrane (*Millipore*) using upward capillary transfer method. This was done in 10X SSC buffer (1.5 M NaCl, 150 mM Na3C6H5O7) and took around 7-8 hours for the transfer of DNA onto the membrane. The DNA on the membrane was then UV crosslinked (*Stratalinker*). The membrane was incubated with 30 ml of pre-hybridization solution (9 ml of 20xSSC, 3ml of 50xDen, 1.5 ml of 10%SDS, 600 µl of 10 µg/µl denatured carrier DNA and 16ml MQ) in rolling tubes at 67°C for an hour.

For hybridization probe synthesis, 150 ng of linearized and gel purified HPV 18 genome in 15µl MQ was used to which 5 µl of deca label oligos were added, mixed, and denatured at 100°C for 5 minutes. The mixture was cooled on ice for a minute, then 1.5 µl Mix C (Mixture of all dNTPs except dCTP), 3 µl of α-32P-dCTP isotope (*Hartman analytics*) and 0.6 µl of Klenow fragment were added and mixed. This mixture was incubated at 37°C for 15 minutes and then 2 µl of dNTPs were added, mixed, and again incubated at 37°C for 15 minutes. Then the mixture was denatured at 100°C for 5 minutes. This mixture was added into the pre-hybridization solution and poured back into the rolling tube which contains the nylon membrane. The tube was incubated at 67°C overnight. Next day the hybridization solution

was discarded, and the membrane was washed with solution I, II, and III. Solution I (10x SSC, 0.1% SDS) was kept for 10 minutes for washing and was repeated; solution II (5x SSC, 0.1% SDS) was kept for 15 minute and solution III (0.5x SSC, 0.1% SDS) was kept for 15 minutes and repeated. The solution was discarded, and the membrane was exposed to phosphorimager screens in cassettes overnight. Then the filter was imaged using phosphorimager (Typhoon Trio, GE Healthcare). The image was later converted into quantified data using ImageQuant software to get the signal values for each sample.

3.1.10 Luciferase assay

Media from 96-well plates used for luciferase assays was aspirated. Substrate mixture containing 95µl of media, 5 µl of MTT reagent (3-[4,5-dimethylthiazol-2yl]-2,5-diphenyl-tetrazolium bromide), 0.125 µl of luciferase assay substrate was added into each well using a multichannel pipette. The plates were then immediately measured for chemiluminescence using GloMax 96 microplate Luminometer (Promega). After 30 minutes the MTT values were measured at 490/700 nm using a colorimeter, to normalize the Nluc activity.

3.1.11 Total RNA extraction, cDNA synthesis and qPCR

Total RNA was extracted from transfected U2OS cells using Quick RNA MiniPrep Kit (*Zymo Research)* according to the manufacturer's protocol. 4-6 µg of eluted RNA was treated with 1.5 µl Turbo DNase (*Thermo Fisher Scientific)* for 1,5 hours at 37°C and later DNase was inactivated at 75°C for 10 minutes, after adding 1 µl of 0.5M EDTA. The RNA samples were then precipitated using 2.5M LiCl, washed with 75% cold ethanol and resuspended in 15 µl MQ water.

For cDNA synthesis, RevertAid first strand cDNA synthesis kit (*Thermo fisher scientific*) was used according to the manufacturer's protocol. 1.25 µg reprecipitated RNA was used, to which 1 µl of Oligo-dT primers was added, incubated at 65°C for 5 minutes and cooled down on ice. cDNA synthesis was done by using 4 µl of 5x reaction buffer, 1 µl of RiboLock RNase inhibitor, 1 µl of RevertAid reverse transcriptase and 2 µl of dNTPs and incubating at 43°C for an hour and then terminating the reaction by incubating at 70°C for 5 minutes.

For qPCR, five separate primer pairs were selected (Table 1). GAPDH was used as a control, E2, E7, E8^E2 and E1^E4 were used to observe HPV18 transcripts originating from multiple

promoters. 5x EVAgreen qPCR master-mix (*Solis Biodyne*) was used along with the selected primers. The synthesized cDNA template from each sample was transferred into qPCR plate wells using a multichannel pipette. To the same wells with the cDNA template, the primer and 5x EVAgreen qPCR master mix was added, and the plate was centrifuged at 2500 RPM. The expression profiles of HPV18 transcripts were determined using qPCR (LightCycler 480, Roche) with a program optimized for selected primers.. The quantified data was calculated by using double delta Ct (cycle threshold) method.

Primers	Sequence
GAPDH	Forward- CTCTCTGCTCCTCCTGTTCGAC Reverse- TGAGCGATGTGGCTCGGCT
E2	Forward- GATAGTGGCTATGGCTGTTC Reverse- GCTGTTGTTGCCCTCTGTG
E7	Forward- CATTTACCAGCCCGACGAG Reverse- GGTCGTCTGCTGAGCTTTC
E8^E2	Forward- GATAGTGGCTATGGCTGTTC Reverse- GACGTCTGGCCGTAGGTCTTTGC
E1^E4	Forward-CATTTACCAGCCCGACGAG Reverse-GACGTCTGGCCGTAGGTCTTTGC

Table 1: Primers used for qPCR.

3.1.12 Statistical analysis

To find statistical significance among the mutant samples for the quantified data results from luciferase assay, southern blot, and qPCR, t tests were performed to calculate the p-value. For luciferase assay and Southern blot results, two-tailed paired t-tests were done and between 48h and 96h time-points for each mutant, multiple-unpaired t tests were done. For, qPCR results, two-tailed one sample t test were done. The significance values were marked accordingly on the result graphs. ($p > 0.05$ = n**s** - not significant; $p \leq 0.05$ = * - significant; $p \leq 0.01$ = ** - very significant.)

3.2 RESULTS

3.2.1 Generation of mutant HPV18 genomes

Five different octamer sequences (binding sites) highly similar to consensus ATTTGCAT were identified on the URR of HPV18 (Figure 7, A). The octamer sequences were substituted with non-binding octamer sequence. For each five mutants, the insert sequence for cloning had one of binding sites altered and the sixth mutant had all binding sites replaced (Figure 7, B).

The mutant sequences were cloned into pMC-HPV18 plasmids and pMC-HPV18 Nluc plasmids and verified using restriction analysis and DNA sequencing. The generated mutants were confirmed by sequencing.

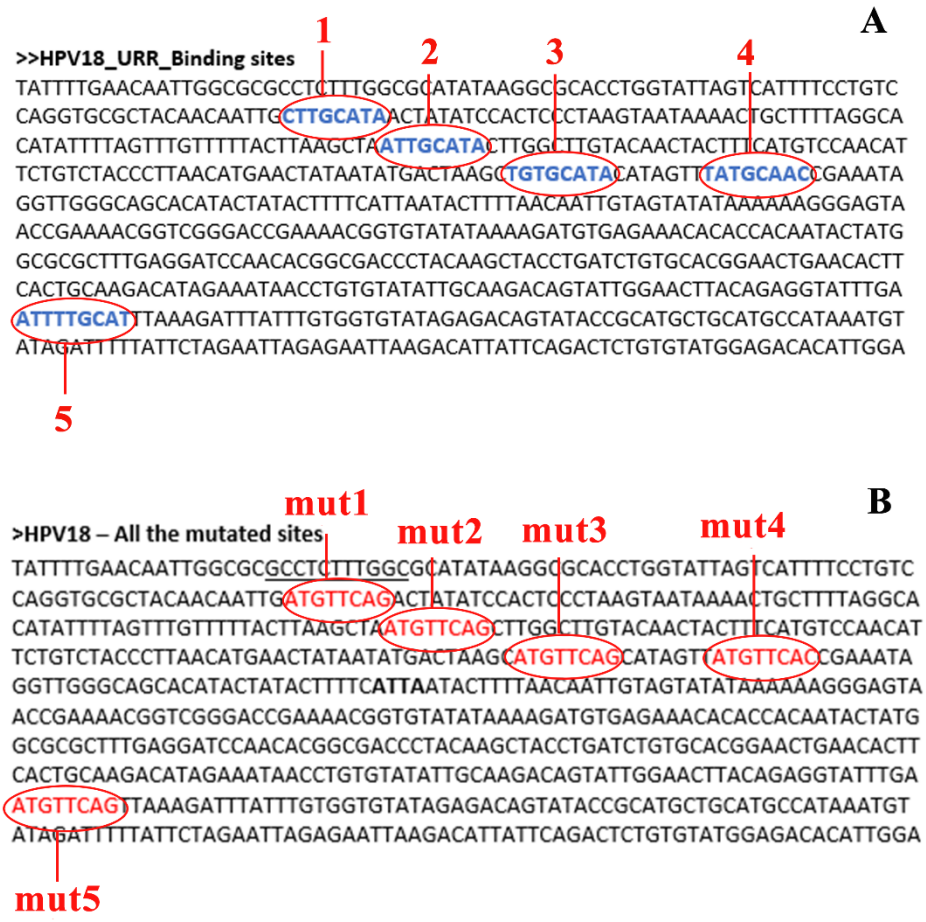

Figure 7: **A**. The putative binding sites identified in the HPV18 URR using bioinformatic analysis; **B**. Depiction of the mutations introduced into the putative Oct-1 binding sites.

3.2.2 Generation of minicircles (MC) from cloned pMC plasmids.

The confirmed mutants cloned into pMC-HPV18 plasmids were converted into minicircles (MC) as described in the methods section. The MCs generated from each mutant were digested using Bgl I, EcoR I and BamH I enzymes and separated on gel by electrophoresis, this helped to determine and confirm the formation of minicircles from their pMC plasmid backbones (Figure 8). The pMC- HPV18 plasmid length is 11947 bps and the minicircles generated from these plasmids are 7948 bps long (Toots et al., 2014). Both the pMC and MC are linearized when restricted with Bgl I, and it can be seen that the MCs are just around the 8kb. Both pMCs and MCs were double cut by the enzyme EcoR I, but for the MC the first cut is shorter compared to pMCs. BamH I introduces double cuts in the pMC, while it only linearizes the MC. Although MCs were formed, the samples had still some non-recombined pMC remaining, and it can be seen as very light bands on the lanes of restricted MC samples. Eventually all the samples for both WT and Nluc mutants were converted into minicircles and restricted to confirm the same. The formed minicircles (MC) of the mutated HPV 18 pMC plasmids were used for replication assays. The MC mutants was used for southern blot analysis and MC Nluc mutants were used for luciferase assay.

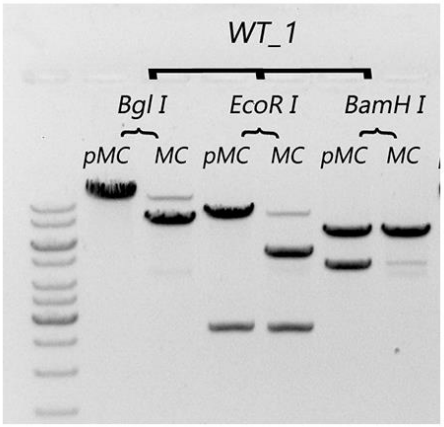

Figure 8: Restriction analysis of the mutants as pMC plasmids and minicircle samples

3.2.3 Transfection efficiency in U2OS cells

Transfection of U2OS cells with the GFP expressing plasmid showed the efficiency of the transfection that is achieved when transfecting the cells with mutant genomes. The efficiency was measured using FACS analysis. The values (Figure 9) indicates that the transfection efficiency was around 90% from the total cells transfected.

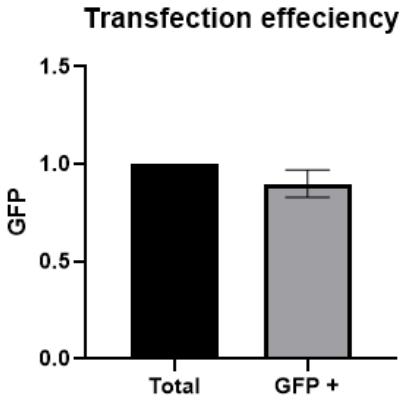

Figure 9: GFP signal gathered from U2OS cells transfected by GFP expression plasmid.

3.2.4 Effect of Oct-1 overexpression on HPV18 replication

To observe the effect of Oct-1 overexpression on the HPV18 replication, the cells were co-transfected with the wild type HPV18 genome and the Oct-1 expression plasmid. As a control pCDNA3 was used as an empty vector in co-transfection.

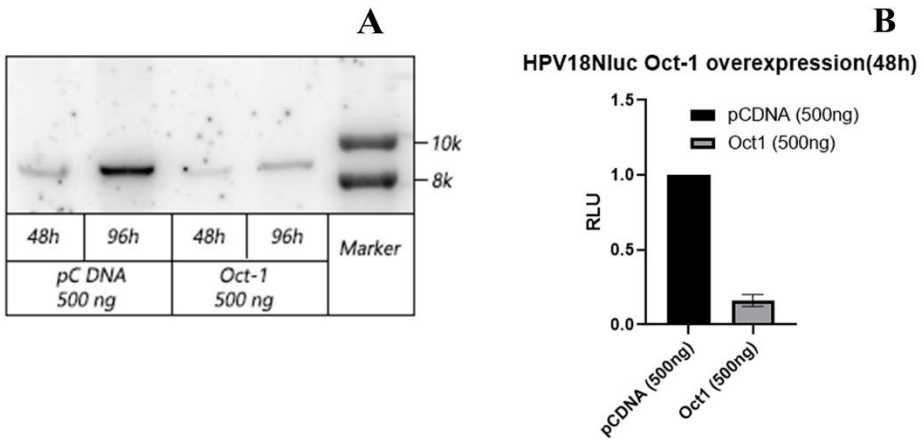

Figure 10: **A**. The effect of overexpression of Oct-1 on the HPV18 genome replication efficiency as analysed by Southern blot a; **B**. Effect of Oct-1 overexpression analysed by luciferase assay.

Southern blot analysis of the genomic DNA extracted from transfected cells showed that overexpression of Oct-1 inhibited HPV18 genome replication at both timepoints (48 hours and 96 hours after transfection) as compared to the samples where HPV18 genome was co-transfected with pCDNA. This result clearly shows that Oct-1 overexpression has a negative effect on HPV18 replication (Figure 10, A). Luciferase assay results were in concert with the Southern blot data and showed similar inhibitory effect of Oct-1 (Figure 10, B).

3.2.5 Effect of OCT-1 binding site mutations on HPV18 replication

Since certain putative binding sites for Oct-1 had been identified, and overexpression of Oct-1 has shown a negative effect on HPV replication, it was hypothesized that mutating these putative binding sites should positively impact the replication efficiency of HPV18 genome.

Replication efficiencies of these Oct-1 binding site mutants were measured by performing luciferase and Southern blot analysis. The observed luciferase signal from each sample co-relates with its replication efficiency in U2OS cells, the relative luciferase activity was measured for each mutant sample by normalizing against the wild-type sample (control).

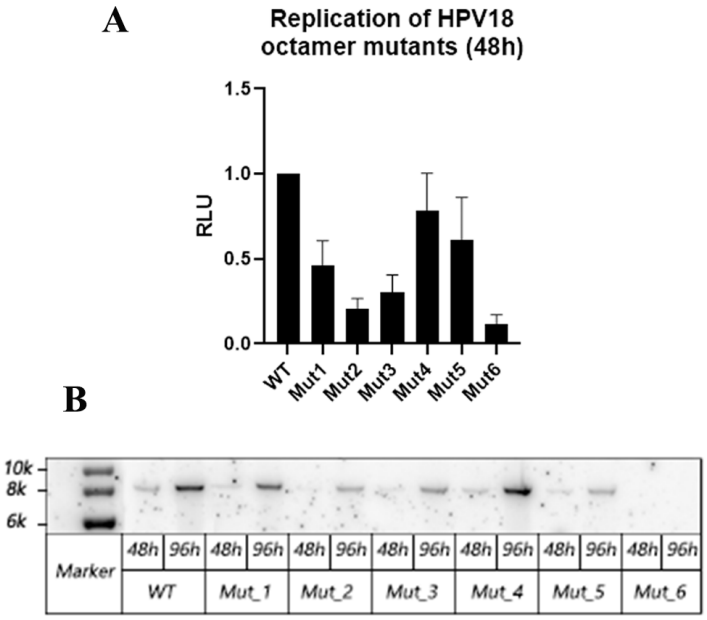

Figure 11: **A**. Varying replication efficiencies of the mutants, measured by relative luminescence units (RLU); **B**. Southern blot image showing the different intensity of signals observed from each of the mutant.

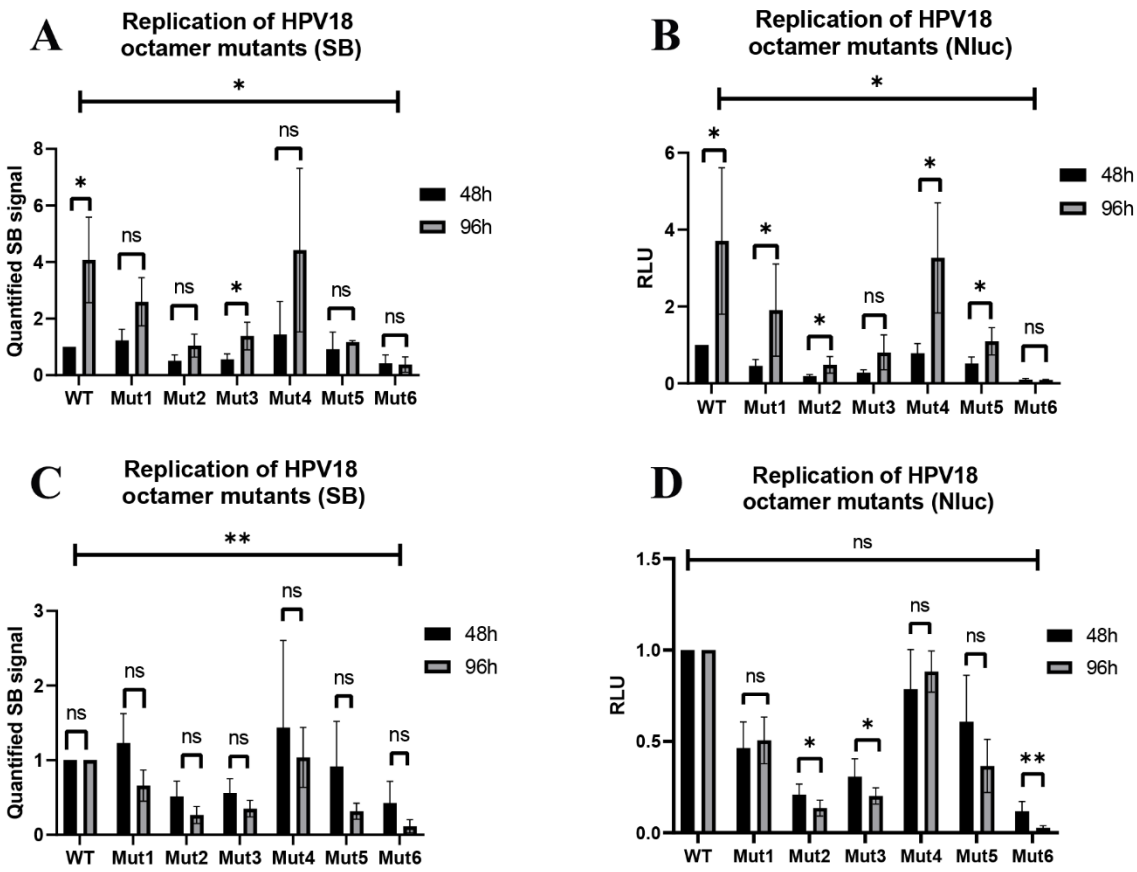

Figure 12. Replication efficiency measured in time, for southern blot and luciferase assays. **A**. Quantified Southern blot data (SB); and **B**. Luciferase analysis (Nluc) show that most of the mutants seem to be replicating in time. Similar effect can be observed for **C**. and **D**., where the data for each timepoint has been shown separately, in comparison to the WT replication signal.

The replication efficiencies for each mutant varied among each other, quite contrasting to how it was expected to perform. Some mutants showed a somewhat neutral effect on the HPV replication efficiency as compared to other mutants which seemed to have a negative effect (Figure 11, A). Southern blot analysis showed a similar pattern of replication efficiencies (Figure 11, B).

Every mutant genome except for the 6th mutant (Mut6) replicated in time. Quantified Southern blot data and Luciferase assay results (Figure 12, A & B) showed that the 1st Mutant, 4th Mutant and the 5th Mutant seemed to be replicating the most, compared to the other mutants. 4th Mutant seemed to enhance the replication efficiencies, as observed in Southern blot and luciferase assays. The same pattern of effect was seen when each mutant genome was normalized against the WT genome (control) (Figure 12 C & D).

3.2.6 Effect of OCT-1 binding site mutations on HPV18 transcription

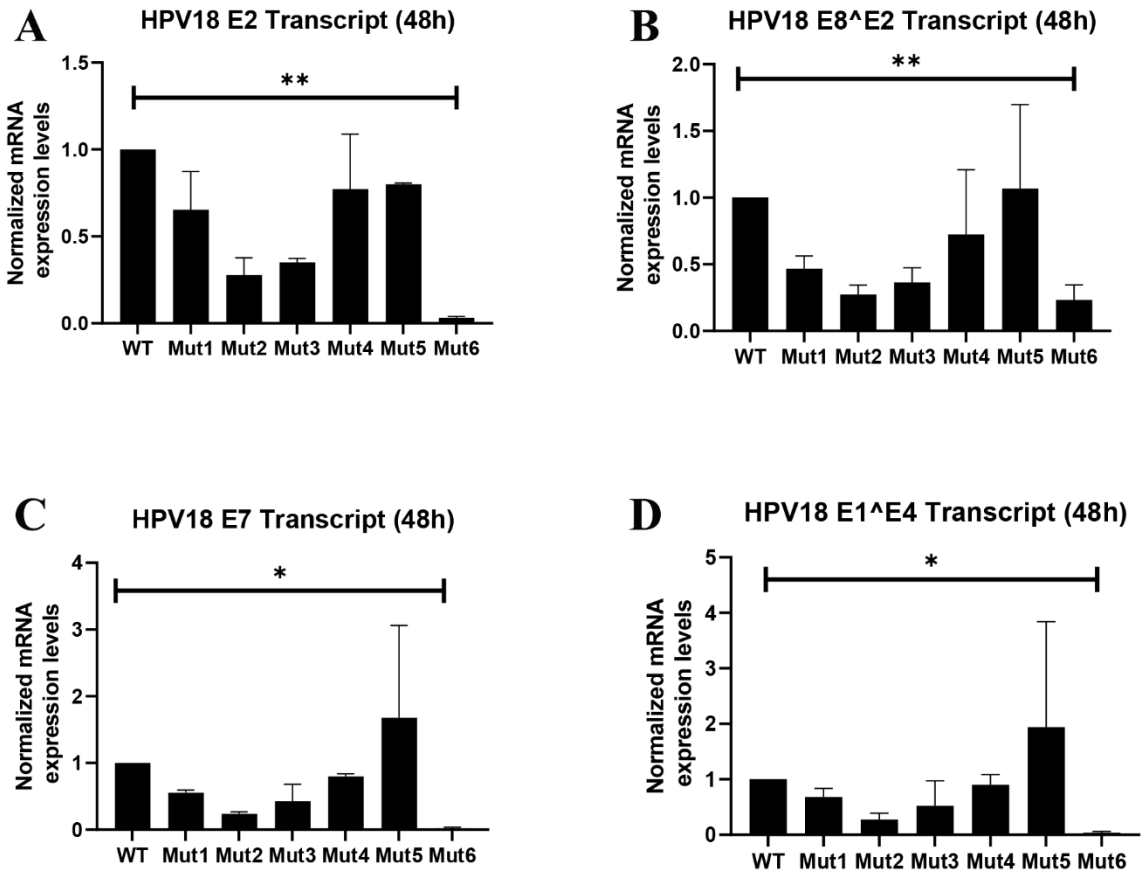

Figure 13: mRNA level expression of different HPV18 transcripts in 48 hours normalized against the WT; **A**. HPV18 E2 transcript; **B**. HPV18 E7 transcript; **C**. HPV18 E8^E2 transcript; **D**. HPV18 E1^E4 transcript

Since, the Oct-1 binding site mutants had some varying effect on the HPV18 replication, it was possible that a similar pattern of effect should be observed for HPV18 mRNA expression levels for different transcripts.

The U2OS cells transfected with the mutant genomes were extracted for total RNA and cDNA was synthesized as mentioned in methods section above. qPCR reactions were prepared with the cDNAs synthesized from each sample transfected with generated mutants' genomes using primer sets for HPV18 E2, E7, E8^E2 and E1^E4 transcripts (Figure 13). The expression of these transcripts seemed to be varying among different mutants. Mutant 6 (Mut6) showed no or very little expression of E2, E7, E8^E2 and E1^E4 transcripts. While other mutants did have some levels of expression, the highest impact was seen to be coming

from Mutants 4 and 5 (Mut4 and Mut5), with the highest expression levels showed by Mutant 5 and for E7 transcript, even much more than the WT control (Figure 13, B). The expression of E2 transcript seemed to be the highest in Mutant 4 and Mutant 5 (Figure 13, A).

3.2 DISCUSSION

The need for studying different aspects of the HPV viral infection comes from the severity of the infections caused certain HPV types. Studies are conducted in order to understand the role of host cell factors involved in the HPV genome replication and transcription, and how these cellular factors could be potentially used against the HPV infection cycle.

The reason OCT-1 was chosen for this study is because, it has been previously shown in our research group (data unpublished) that many octamer family transcriptional factors interact with URR and have also shown to have some effect on HPV replication. Also, OCT-1 is abundantly expressed in U2OS cell lines, which makes it a good candidate to understand the effect of OCT-1 on viral replication. Oct-1 is known to bind to certain motifs along the URR of the HPV genome, and bioinformatics analysis have suggested certain putative binding sites for Oct-1. Our research group has shown that the overexpression of OCT-1 has a significant effect on HPV18 replication, so it was hypothesized that by mutational analysis of these putative binding sites, it could be investigated which Oct-1 binding site has the most impact.

This study achieved that, by constructing mutants of the predicted Oct-1 binding sites and perform replication assays and transcriptional analysis to find significance of these binding sites on HPV 18 genome replication as well as transcription.

The results obtained from Southern blot analysis and luciferase assay, backed by transcriptional analysis, showed that each binding site had a different impact on the HPV 18 replication and transcription. Some binding site mutants showed to have a negative impact on the HPV replication such as the 2nd and 3rd mutant; This possibly means either that the sites were important for HPV replication or the site was overlapping with other important factors that may be crucial for genome replication. In case of the 6th mutant, altering out all the putative binding sites seemed to make the viral genome not replicate at all. Interestingly enough; 1st mutant, 4th mutant and 5th mutant seemed to have lesser impact towards the replication effi-

ciencies than other mutants. Furthermore, the 4th mutant (Mut4) showed an increased replication efficiency (Figure 11, A and C), which could mean that the fourth binding site maybe crucial for Oct-1 induced negative effect on HPV18 replication.

Observing the genome map and organization of these putative Oct-1 binding sites found in contrast to the E1 and E2 binding sites (Figure 14), a hypothesis could be derived as to why certain mutants could have a positive effect on the HPV replication.

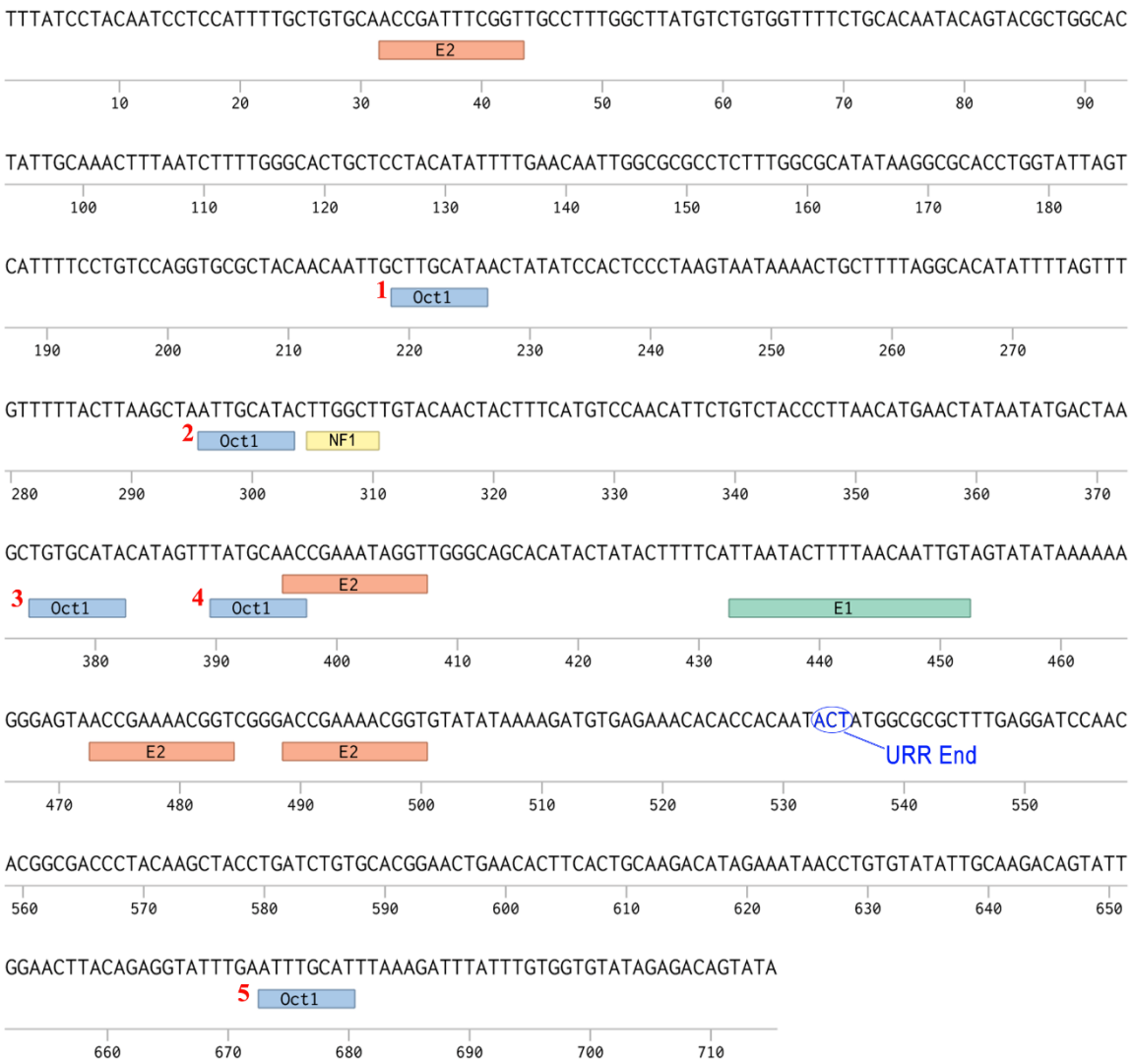

Figure 14: The genome map indicating the presence of the putative Oct-1 binding sites (labelled in red numbers), NF1 binding site, location of E1 and E2 binding sites and the URR end.

The 4th putative binding site for Oct-1, lies next to the E2 binding site on the URR, and it also overlaps the E2 binding site by 2 nucleotides. The mutant constructed for this site was designed in such a way that the last 2 overlapping nucleotides remain unaltered (last G was

substituted by C) and hence not mutating the E2 binding site as a result. Since, E2 binding is important for HPV replication, and the 4th binding site being so close to the E2 binding site, Oct-1 may compete with E2 and hence negatively impact replication. Hence, mutation to the 4th binding site seems to have a positive effect on HPV replication. However, more experiments are needed to form concrete conclusions about these binding sites.

The 2nd Oct-1 binding site is just adjacent to the NF1 binding site (Morris et al., 1993), and this maybe the reason that the 2nd mutant had a really lowered replication efficiency. One possible hypothesis for this effect could be that Oct-1 and NF-1 may work synergistically for the activation of the URR (Chong et al., 1991).

The 5th binding site is complementary to the actual octamer consensus (ATGCAAAT) and it lies beyond the URR, on the E6 open reading frame. From transcriptional analysis result, it could be hypothesized that the 5th mutant has a positive impact on expression of different mRNA transcripts. This might be either due to increased accessibility on the region or conformational changes (or combined). Although, this needs to be further analysed, to understand, what exactly is the mechanism behind this effect, hence more experiments are needed.

The cumulative understanding gained from previous studies on potential effect of Oct-1 transcriptional factors on HPV18 replication and the set of experiments done in this thesis study to investigate the impact of these putative Oct-1 binding sites on HPV18 replication could be important to understand the role of certain cellular factors in HPV life cycle and to develop novel strategies to inhibit the viral infection.

SUMMARY

The focus of this thesis was on investigating the effect of putative Oct-1 binding sites on HPV18 replication and observe their impact on expression of HPV18 transcripts. This was achieved by constructing mutant HPV18 genomes with the putative Oct-1 binding sites altered with non-binding sequences. Each HPV18 mutant genome was transfected into U2OS cells to observe their replication efficiency and gene expression pattern. Each mutant had varying effects, some mutants stood out with their effect on HPV18 replication and transcription. The logic behind these effects were hypothesized and discussed for possible mechanisms through which Oct-1 could induce a negative impact on HPV18 replication and transcription:

- 4^{th} binding site overlaps the E2 binding site by 2 nucleotides, hence it could possibly compete with E2 viral protein.
- 2^{nd} binding site lies just close to the NF1 binding site, which could be the reason for lower replication and transcription.
- The role of the 5^{th} binding site needs to be understood more clearly, since mutation to that site seems to increase the expression of different HPV18 transcripts.
- All the theory discussed behind such effects are still hypothetical and hence they need to be further confirmed by conducting more experiments.

In conclusion, the identified putative Oct-1 binding sites do seem to have some potential mechanisms by which it brings about an inhibitory effect on HPV18 genome replication and transcription, and the exact mechanisms behind these effects are still left to be understood.

REFERENCES

Aksoy, P., Gottschalk, E. Y., & Meneses, P. I. (2017). HPV entry into cells. In *Mutation Research - Reviews in Mutation Research* (Vol. 772, pp. 13–22). Elsevier B.V. https://doi.org/10.1016/j.mrrev.2016.09.004

Archambault, J. (2011). Small Molecule Inhibitors of Human Papillomavirus Protein-Protein Interactions. In *The Open Virology Journal* (Vol. 5).

Bauknecht, T., See, R. H., & Shi, Y. (1996). A novel C/EBP beta-YY1 complex controls the cell-type-specific activity of the human papillomavirus type 18 upstream regulatory region. *Journal of Virology*, *70*(11), 7695–7705. https://doi.org/10.1128/jvi.70.11.7695-7705.1996

Boner, W., Taylor, E. R., Tsirimonaki, E., Yamane, K., Saveria Campo, M., & Morgan, I. M. (2002). *A Functional Interaction between the Human Papillomavirus 16 Transcription/Replication Factor E2 and the DNA Damage Response Protein TopBP1**. https://doi.org/10.1074/jbc.M202163200

Breiding, D. E., Sverdrup, F., Grossel, M. J., Moscufo, N., Boonchai, W., & Androphy, E. J. (1997). Functional interaction of a novel cellular protein with the papillomavirus E2 transactivation domain. *Molecular and Cellular Biology*, *17*(12), 7208–7219. https://doi.org/10.1128/mcb.17.12.7208

Bromberg-White, J. L., & Meyers, C. (2002). The Upstream Regulatory Region of Human Papillomavirus Type 31 Is Insensitive to Glucocorticoid Induction. *JOURNAL OF VIROLOGY*, *76*(19), 9702–9715. https://doi.org/10.1128/JVI.76.19.9702-9715.2002

Burley, M., Roberts, S., & Parish, J. L. (2019). *Epigenetic regulation of human papillomavirus transcription in the productive virus life cycle*. https://doi.org/10.1007/s00281-019-00773-0/Published

Carson, A., & Khan, S. A. (2006). Characterization of Transcription Factor Binding to Human Papillomavirus Type 16 DNA during Cellular Differentiation. *JOURNAL OF VIROLOGY*, *80*(9), 4356–4362. https://doi.org/10.1128/JVI.80.9.4356-4362.2006

Chan, W. K., Chong, T., Bernard, H. U., & Kiock, G. (1990). Transcription of the transforming genes of the oncogenic human papillomavirus-16 is stimulated by tumor promotors

through AP1 binding sites. *Nucleic Acids Research*, *18*(4), 763–769. https://doi.org/10.1093/nar/18.4.763

Chong, T., Apt, D., Gloss, B., Isa, M., & Bernard, H.-U. (1991). The Enhancer of Human Papillomavirus Type 16: Binding Sites for the Ubiquitous Transcription Factors oct-1, NFA, TEF-2, NF1, and AP-1 Participate in Epithelial Cell-Specific Transcription Downloaded from. In *JOURNAL OF VIROLOGY* (Vol. 65, Issue 11). http://jvi.asm.org/

Chong, T., Chan, W. K., & Bernard, H. U. (1990). Transcriptional activation of human papillomavirus 16 by nuclear factor I, AP1, steroid receptors and a possibly novel transcription factor, PVF: A model for the composition of genital papillomavirus enhancers. *Nucleic Acids Research*, *18*(3), 465–470. https://doi.org/10.1093/nar/18.3.465

de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U., & zur Hausen, H. (2004). Classification of papillomaviruses. In *Virology* (Vol. 324, Issue 1, pp. 17–27). Academic Press. https://doi.org/10.1016/j.virol.2004.03.033

Demeret, C., le Moal, M., Yaniv, M., & Thierry, F. (1995). Control of HPV 18 DNA replication by cellular and viral transcription factors. In *Nucleic Acids Research* (Vol. 23, Issue 23).

Dong, X. -P, Stubenrauch, F., Beyer-Finkler, E., & Pfister, H. (1994). Prevalence of deletions of YY1-binding sites in episomal HPV 16 DNA from cervical cancers. *International Journal of Cancer*, *58*(6), 803–808. https://doi.org/10.1002/ijc.2910580609

Doorbar, J., Quint, W., Banks, L., Bravo, I. G., Stoler, M., Broker, T. R., & Stanley, M. A. (2012). The biology and life-cycle of human papillomaviruses. In *Vaccine* (Vol. 30, Issue SUPPL.5, pp. F55–F70). Elsevier Ltd. https://doi.org/10.1016/j.vaccine.2012.06.083

Fisher, C. (2015). Recent Insights into the Control of Human Papillomavirus (HPV) Genome Stability, Loss, and Degradation. *Journal of Clinical Medicine*, *4*(2), 204–230. https://doi.org/10.3390/jcm4020204

Geimanen, J., Isok-Paas, H., Pipitch, R., Salk, K., Laos, T., Orav, M., Reinson, T., Ustav, M., & Ustav, E. (2011). Development of a Cellular Assay System To Study the Genome Replication of High-and Low-Risk Mucosal and Cutaneous Human Papillomaviruses Downloaded from. *JOURNAL OF VIROLOGY*, *85*(7), 3315–3329. https://doi.org/10.1128/JVI.01985-10

Gillison, M. L., Chaturvedi, A. K., & Lowy, D. R. (2008). HPV prophylactic vaccines and the potential prevention of noncervical cancers in both men and women. In *Cancer* (Vol. 113, Issue 10 SUPPL., pp. 3036–3046). NIH Public Access. https://doi.org/10.1002/cncr.23764

Gloss, B., & Bernard, H. U. (1990). The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by a sequence-aberrant Sp1 distal element. *Journal of Virology*, *64*(11), 5577–5584. https://doi.org/10.1128/jvi.64.11.5577-5584.1990

Gold, D. A., Gates, R. D., & Jacobs, D. K. (2014). The Early Expansion and Evolutionary Dynamics of POU Class Genes. *Molecular Biology and Evolution*, *31*(12), 3136–3147. https://doi.org/10.1093/molbev/msu243

Graham, S. v. (2010). Human papillomavirus: Gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. In *Future Microbiology* (Vol. 5, Issue 10, pp. 1493–1506). Europe PMC Funders. https://doi.org/10.2217/fmb.10.107

Graham, S. v. (2017a). The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. *Clinical Science*, *131*, 2201–2221. https://doi.org/10.1042/CS20160786

Graham, S. v. (2017b). Keratinocyte differentiation-dependent human papillomavirus gene regulation. In *Viruses* (Vol. 9, Issue 9). MDPI AG. https://doi.org/10.3390/v9090245

Hoffmann, R., Hirt, B., Bechtold, V., Beard, P., & Raj, K. (2006). Different Modes of Human Papillomavirus DNA Replication during Maintenance. *Journal of Virology*, *80*(9), 4431–4439. https://doi.org/10.1128/jvi.80.9.4431-4439.2006

Kadaja, M., Silla, T., Ustav, E., & Ustav, M. (2009). Papillomavirus DNA replication - From initiation to genomic instability. In *Virology* (Vol. 384, Issue 2, pp. 360–368). Academic Press. https://doi.org/10.1016/j.virol.2008.11.032

Kitagawa, K., Yoshikawa, H., Onda, T., Kawana, T., Taketani, Y., Yoshikura, H., & Iwamoto, A. (1996). Genomic organization of human papillomavirus type 18 in cervical cancer specimens. *Japanese Journal of Cancer Research*, *87*(3), 263–268. https://doi.org/10.1111/j.1349-7006.1996.tb00215.x

Konzak, K. E., & Moore, D. D. (1992). *Functional lsoforms of Pit-l Generated by Alternative Messenger RNA Splicing*. https://academic.oup.com/mend/article/6/2/241/2897221

Kyo, S., Inoue, M., Nishio, Y., Nakanishi, K., Akira, S., Inoue, H., Yutsudo, M., Tanizawa, O., & Hakura, A. (1993). NF-IL6 represses early gene expression of human papillomavirus type 16 through binding to the noncoding region. *Journal of Virology*, *67*(2), 1058–1066. https://doi.org/10.1128/jvi.67.2.1058-1066.1993

Latchman, D. S. (1999). Regulation of DNA Virus Transcription by Cellular POU Family Transcription Factors. In *Reviews in Medical Virology Rev. Med. Virol* (Vol. 9).

Lee, D., Hwang, S. G., Kim, J., & Choe, J. (2002). Functional interaction between p/CAF and human papillomavirus E2 protein. *Journal of Biological Chemistry*, *277*(8), 6483–6489. https://doi.org/10.1074/jbc.M105085200

Lee, D., Lee, B., Kim, J., Kim, D. W., & Choe, J. (2000). cAMP response element-binding protein-binding protein binds to human papillomavirus E2 protein and activates E2-dependent transcription. *Journal of Biological Chemistry*, *275*(10), 7045–7051. https://doi.org/10.1074/jbc.275.10.7045

Liu, Y., Lu, Z., Xu, R., & Ke, Y. (2015). *Comprehensive mapping of the human papillomavirus (HPV) DNA integration sites in cervical carcinomas by HPV capture technology* (Vol. 7, Issue 5). www.impactjournals.com/oncotarget

Ma, T., Zou, N., Lin, B. Y., Chow, L. T., & Harper, J. W. (1999). Interaction between cyclin-dependent kinases and human papillomavirus replication-initiation protein E1 is required for efficient viral replication. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(2), 382–387. https://doi.org/10.1073/pnas.96.2.382

Malik, H., Khan, F. H., & Ahsan, H. (2014). Human papillomavirus: Current status and issues of vaccination. In *Archives of Virology* (Vol. 159, Issue 2, pp. 199–205). https://doi.org/10.1007/s00705-013-1827-z

Miller Rank, N., & Lambert, P. F. (1995). Bovine Papillomavirus Type 1 E2 Transcriptional Regulators Directly Bind Two Cellular Transcription Factors, TFIID and TFIIB. In *JOURNAL OF VIROLOGY* (Vol. 69, Issue 10). http://jvi.asm.org/

Moody, C. A. (2017). *Mechanisms by which HPV Induces a Replication Competent Environment in Differentiating Keratinocytes*. https://doi.org/10.3390/v9090261

Morris, P. J., Dent+, C. L., Ring, C. J., & Latchman, D. S. (1993). The octamer binding site in the HPV16 regulatory region produces opposite effects on gene expression in cervical and non-cervical cells. In *Nucleic Acids Research* (Vol. 21, Issue 4).

O'Connor, M., & Bernard, H. U. (1995). Oct-1 activates the epithelial-specific enhancer of human papillomavirus type 16 via a synergistic interaction with NFI at a conserved composite regulatory element. *Virology*, *207*(1), 77–88. https://doi.org/10.1006/viro.1995.1053

O'connor, M., Chan, S., & Bernard, H. (1995). *Transcription Factor Binding Sites in the Long Control Region of Genital HPVs*.

Orav, M., Geimanen, J., Sepp, E.-M., Henno, L., Ustav, E., & Ustav, M. (2015). *Initial amplification of the HPV18 genome proceeds via two distinct replication mechanisms OPEN*. https://doi.org/10.1038/srep15952

Parkin, D. M. (2006). *The global health burden of infection-associated cancers in the year 2002*. https://doi.org/10.1002/ijc.21731

Phillips, K., & Luisi, B. (2000). The virtuoso of versatility: POU proteins that flex to fit. In *Journal of Molecular Biology* (Vol. 302, Issue 5, pp. 1023–1039). Academic Press. https://doi.org/10.1006/jmbi.2000.4107

Piirsoo, A., Piirsoo, M., Kala, M., Sankovski, E., Lototskaja, E., Levin, V., Salvi, M., & Ustav, M. (2019). Activity of CK2α protein kinase is required for efficient replication of some HPV types. *PLOS Pathogens*, *15*(5), e1007788. https://doi.org/10.1371/journal.ppat.1007788

Rehtanz, M., Schmidt, H.-M., Warthorst, U., & Steger, G. (2004). Direct Interaction between Nucleosome Assembly Protein 1 and the Papillomavirus E2 Proteins Involved in Activation of Transcription. *Molecular and Cellular Biology*, *24*(5), 2153–2168. https://doi.org/10.1128/mcb.24.5.2153-2168.2004

Reményi, A., Tomilin, A., Pohl, E., Lins, K., Philippsen, A., Reinbold, R., Schöler, H. R., & Wilmanns, M. (2001). Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. *Molecular Cell*, *8*(3), 569–580. https://doi.org/10.1016/S1097-2765(01)00336-7

Ribeiro, A. L., Caodaglio, A. S., & Sichero, L. (2018). Regulation of HPV transcription. In *Clinics* (Vol. 73, Issue Suppl 1). Universidade de Sao Paulo. https://doi.org/10.6061/clinics/2018/e486s

Saraiya, M., Unger, E. R., Thompson, T. D., Lynch, C. F., Hernandez, B. Y., Lyu, C. W., Steinau, M., Watson, M., Wilkinson, E. J., Hopenhayn, C., Copeland, G., Cozen, W.,

Peters, E. S., Huang, Y., Saber, M. S., Altekruse, S., & Goodman, M. T. (2015). US assessment of HPV Types in cancers: Implications for current and 9-valent HPV vaccines. *Journal of the National Cancer Institute*, *107*(6), 86. https://doi.org/10.1093/jnci/djv086

Scully, K. H., Jacobson, E. M., Jepsen, K., Lunyak, V., Viadiu, H., Carriere, C., Rose, D. W., Hooshmand, F., Aggarwal, A. K., & Rosenfeld, M. G. (2000). Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. In *Science* (Vol. 290, Issue 5494, pp. 1127–1131). Science. https://doi.org/10.1126/science.290.5494.1127

Stu¨nkel, W., & Bernard, H.-U. (1999). The Chromatin Structure of the Long Control Region of Human Papillomavirus Type 16 Represses Viral Oncoprotein Expression Downloaded from. In *JOURNAL OF VIROLOGY* (Vol. 73, Issue 3). http://jvi.asm.org/

Sturm, R. A., Das, G., & Herr, W. (1988). *The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain*.

Sverdrup, F., & Khan, S. A. (1994). Replication of Human Papillomavirus (HPV) DNAs Supported by the HPV Type 18 El and E2 Proteins Downloaded from. In *JOURNAL OF VIROLOGY* (Vol. 68). http://jvi.asm.org/

Sverdrup, F., & Khan, S. A. (1995). Two E2 Binding Sites Alone Are Sufficient To Function as the Minimal Origin of Replication of Human Papillomavirus Type 18 DNA Downloaded from. In *JOURNAL OF VIROLOGY* (Vol. 69, Issue 2). http://jvi.asm.org/

Toots, M., Männik, A., Kivi, G., Ustav, M., Ustav, E., & Ustav, M. (2014). The transcription map of human papillomavirus type 18 during genome replication in U2OS cells. *PLoS ONE*, *9*(12), e116151. https://doi.org/10.1371/journal.pone.0116151

van Doorslaer, K., Chen, D., Chapman, S., Khan, J., & Mcbride, A. A. (2017). *Persistence of an Oncogenic Papillomavirus Genome Requires cis Elements from the Viral Transcriptional Enhancer*. https://pave.niaid.nih.gov]

van Doorslaer, K., Chen, Z., Bernard, H.-U., Chan, P. K. S., Desalle, R., Dillner, J., Forslund, O., Haga, T., Mcbride, A. A., Villa, L. L., & Burk, R. D. (2018). *ICTV Virus Taxonomy Profile: Papillomaviridae*. https://doi.org/10.1099/jgv.0.001105

Vázquez-Arreguín, K., & Tantin, D. (2016). The Oct1 transcription factor and epithelial malignancies: Old protein learns new tricks. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1859*(6), 792–804. https://doi.org/10.1016/j.bbagrm.2016.02.007

Wang, X., Meyers, C., Wang, H.-K., Chow, L. T., & Zheng, Z.-M. (2011). Construction of a Full Transcription Map of Human Papillomavirus Type 18 during Productive Viral Infection §. *JOURNAL OF VIROLOGY*, *85*(16), 8080–8092. https://doi.org/10.1128/JVI.00670-11

Williams, V. M., Filippova, M., Soto, U., & Duerksen-Hughes, P. J. (2011). HPV-DNA integration and carcinogenesis: Putative roles for inflammation and oxidative stress. In *Future Virology* (Vol. 6, Issue 1, pp. 45–57). NIH Public Access. https://doi.org/10.2217/fvl.10.73

Yu, J.-H., Lin, Y., Deng, W., Broker, T. R., & Chow, L. T. (2007). Mitogen-Activated Protein Kinases Activate the Nuclear Localization Sequence of Human Papillomavirus Type 11 E1 DNA Helicase To Promote Efficient Nuclear Import Downloaded from. *JOURNAL OF VIROLOGY*, *81*(10), 5066–5078. https://doi.org/10.1128/JVI.02480-06

NON-EXCLUSIVE LICENCE TO REPRODUCE THESIS AND MAKE THESIS PUBLIC

I,

Michael Joseph,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

The effect of binding sites for POU-HD proteins (Oct-1) on the replication of HPV18,

supervised by *Marko Piirsoo* and *Martin Kala.*

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Michael Joseph

20/05/2021